



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

METODOLOGI PENELITIAN

3.1 Objek Penelitian

Penelitian ini dilakukan dengan menggunakan data utang pemerintah dalam negeri yang didapatkan dari *kaggle* (Kaggle, 2017), yaitu data yang berisikan besaran uang yang digunakan oleh pemerintah Indonesia pada masing-masing wilayah di Indonesia. Data yang didapatkan ini termasuk data sekunder karena didapatkan melalui media atau perantara yakni melalui website *kaggle*.

Kaggle adalah website resmi yang menyediakan berbagai macam jenis data dari berbagai macam negara. *Kaggle* juga dapat dikatakan sebagai suatu komunitas terbesar yang berisikan para ilmuwan data (*data scientist*) di seluruh dunia. Dalam *kaggle* inilah para ilmuwan data saling berkompetisi untuk menyelesaikan permasalahan ilmiah berdasarkan data yang kompleks atau sering disebut sebagai *big data*. Kompetisi yang diadakan *kaggle* ini juga menyediakan *fee* atau uang untuk penghargaan. Selain itu biasanya akan banyak sekali tawaran bagi pemenangnya untuk bergabung dengan perusahaan-perusahaan besar untuk menyelesaikan masalah bisnis menarik dan sensitif dari perusahaan-perusahaan dunia terbesar.

Melalui *kaggle* inilah banyak sekali data-data yang terbuka yang sifatnya untuk pembelajaran baik pemula maupun yang sudah ahli dalam bidangnya untuk menyelesaikan masalah yang ada. Tidak hanya data

perusahaan tetapi segala macam data dari *datasets* untuk pembelajaran sampai *datasets* milik pemerintah yang sifatnya terbuka untuk membantu negara tersebut.

3.2 Metode Penelitian

Metode-metode yang dimiliki untuk mengolah data menjadi suatu proses data visualisasi dapat menggunakan metode yang sudah disiapkan agar dapat mempermudah analisis dan visualisasi. Metode-metode yang ada yakni KDD, CRISP-DM, dan SEMMA. Metode-metode ini memiliki kelebihan dan kekurangan masing-masing yang dapat dilihat pada gambar 3.1 ini.

STEP / Proses	KDD	Crisp-DM	SEMMA
STEP / PROSES	Developing and Understanding the application	Business understanding	...
	Creating a target datasets	Data understanding	Sample
	Data cleaning and pre-processing		
	Data transformation	Data preparation	Explore
	Choosing the suitable data mining task	Modeling	Modify
	Choosing the suitable data mining algorithm		Model
	Employing data mining algorithm		
	Interpreting mined patterns	Evaluation	Assessment
	Using discovered knowledge	Deployment	...
Total step / proses	9	6	5

Gambar 3.1 Perbandingan metode KDD, CRISP-DM, SEMMA

Pada gambar 3.1 dapat terlihat jelas perbandingan mulai dari tahapan-tahapan metode tersebut mana yang memiliki tahap terbanyak dan terkecil. Jumlah step yang banyak maupun jumlah *step* atau proses yang sedikit tidak

menentukan apakah metode tersebut yang terbaik atau terburuk. Kembali kita harus melihat segala kekurangan dan kelebihan dari masing-masing metode yang ada.

Metode KDD dapat dikatakan cukup lengkap dan kompleks serta dapat dikatakan cukup banyak. Tetapi banyak hal yang ada dalam metode ini dapat dijadikan satu tahapan atau proses sehingga tidak membuang tenaga untuk mengerjakan tahapan-tahapan yang ada

Metode CRISP-DM dapat dikatakan untuk versi lebih cepat, ringkas, dan tepat dalam mengerjakan beberapa penelitian maupun menyelesaikan suatu kasus. Metode ini tidak memiliki tahapan sebanyak metode KDD karena hanya memiliki 6 metode. Metode ini meringkas semua pekerjaan yang ada di KDD menjadi lebih ringkas lagi dan tepat. Metode ini dapat memangkas waktu pengerjaan dengan lebih cepat seperti pada tahap kedua dan ketiga dapat disatukan dan tahap kelima sampai ketujuh juga dapat disatukan sehingga dapat lebih mempercepat pengerjaan

Metode SEMMA ini adalah metode yang memiliki paling sedikit proses atau tahapan-tahapannya yakni hanya memiliki sebanyak 5 tahap. Walaupun hanya berbeda sedikit dengan metode CRISP-DM tidak membuat metode ini menjadi yang terbaik karena masih banyak yang harus diproses tidak hanya dapat dikerjakan dengan 5 tahap tersebut. pada dasarnya harus dimengerti terlebih dahulu permasalahannya baru dapat dibuat *dashboard* untuk menyelesaikan masalah tersebut. SEMMA dalam hal ini langsung menuju

sample yang dimana akan menghabiskan waktu untuk mengerti data tersebut terlebih dahulu.

Perbandingan tersebut membuat penelitian ini menggunakan metode CRISP-DM untuk melakukan proses data mining terhadap data utang pemerintah Indonesia. Berikut ini adalah enam tahap siklus hidup pengembangan data mining untuk visualisasi data transportasi umum di Indonesia:

- a. *Business Understanding*: Tahap pertama adalah memahami permasalahan hutang yang dimiliki oleh masing-masing wilayah di Indonesia. Permasalahannya yaitu hutang yang dimiliki Indonesia dan perbedaan hutang yang ada di masing-masing wilayah Indonesia.
- b. *Data Understanding*: Pada tahap ini, akan diadakan pengumpulan data yang berada pada *kaggle*. Dikarenakan data yang ada pada *kaggle* ini tidak terkumpul menjadi satu tempat maka akan disorting dan mengumpulkannya menjadi satu. Data-data yang berada pada *kaggle* ada yang berbentuk PDF, Excel, dan Word.
- c. *Data Preparation*: Tahap ini akan didapatkan data yang sudah di sortir sehingga dapat membagi-bagi data tersebut berdasarkan atribut-atribut yang diperlukan untuk mengolah data tersebut. Pada tahap ini dapat disebut juga tahap *Data Cleansing* karena pada tahap ini meliputi pembersihan semua data-data yang *null* (kosong), semua typo, perbedaan format data/*type data* dan semua data-data yang perlu dibersihkan sehingga mempermudah dalam mengolah data-data

tersebut. Seperti yang sudah disebutkan di atas, *kaggle* data-datanya tidak satu tipe tetapi dari berbagai macam tipe data.

- d. *Modeling*: Dalam tahap ini akan ditentukan parameter-parameter yang dibutuhkan lalu menetapkan model yang akan digunakan untuk ke tahap berikutnya agar data tersebut mudah untuk diolah. Tahap ini dapat disebut juga *data filtering* yakni data yang sudah dibersihkan akan dibagi menjadi masing-masing kategori berdasarkan parameter-parameter yang sudah ditentukan.

Parameter-parameter tersebut yang sudah ditentukan tersebut akan diisi oleh data-data yang ada pada masing masing kategori serta jika masih ditemukan data yang *null* (kosong) maka akan kembali lagi ke bagian *data cleansing* untuk dibersihkan kembali agar proses filterisasi data dapat berjalan dengan lancar.

- e. *Evaluation*: Pada tahap ini, akan diadakan pengecekan kembali semua data serta pensortiran data-data tersebut apakah sudah sesuai dengan yang diinginkan serta sudah *clean* (bersih) sehingga dapat diolah dan dibuat visualisasinya.

- f. *Deployment*: Tahap ini adalah tahap yang terakhir yakni semua data sudah bersih dan siap diolah. sesuai dengan kebutuhan lalu mengimplementasikannya melalui aplikasi *powerBI* agar dapat divisualisasikan (visualisasi data) dan dapat dimengerti oleh khalayak banyak.

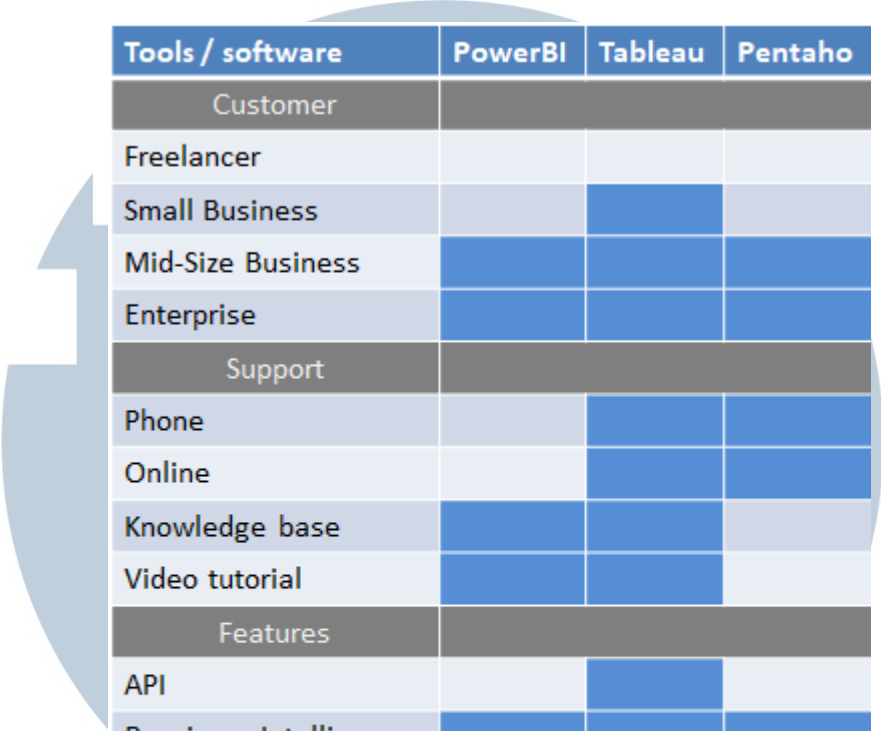
Melalui perbandingan-perbandingan tersebutlah, maka penelitian ini menggunakan metode CRISP-DM sebagai metode utama untuk menganalisis data Utang Pemerintah Dalam Negeri Indonesia dan memvisualisasikannya.

3.3 Metode Analisa Data

Pada tahap ini, penelitian akan menggunakan metode analisa data dengan data kuantitatif. Tahapan-tahapan yang akan dilakukan untuk menganalisis data yaitu:

1. Menyiapkan terlebih dahulu apa saja yang dibutuhkan untuk diolah dalam *powerBI*.
2. Setelah itu menyimpulkan apa saja yang sangat dibutuhkan untuk diolah setelah itu mengolah data-data tersebut dalam *powerBI*.
3. Setelah diolah maka akan mendapat hasil data yang *clean*, setelah itu akan menampilkannya sesuai dengan *chart, bar, pie chart*, dan semua jenis tampilan yang dibutuhkan untuk menunjang hasil yang baik dan mudah dimengerti oleh khalayak banyak.
4. Setelah itu menyimpulkan hasil-hasil yang didapat dari visualisasi data tersebut menggunakan teori analisis data kuantitatif agar dapat mudah dimengerti.

Tools-tools yang paling sering digunakan dan cukup terkenal yakni ada Pentaho, Rapid Miner, serta yang paling terkenal adalah tableau. Masih banyak lagi *tools-tools* yang sering digunakan untuk menganalisis data tetapi masing-masing memiliki kelebihan dan kekurangannya masing-masing. Dibawah ini adalah contoh perbandingan dari masing-masing *tools*.



Tools / software	PowerBI	Tableau	Pentaho
Customer			
Freelancer			
Small Business			
Mid-Size Business			
Enterprise			
Support			
Phone			
Online			
Knowledge base			
Video tutorial			
Features			
API			
Bussiness Intelligence			
Data visualisation			
Dashboard creation			
Ad hoc query			
Customize reporting			

Gambar 3.2 Perbandingan tools

Sumber: [quora.com/Which-tool-is-better-to-learn-Tableau-or-Power-BI](https://www.quora.com/Which-tool-is-better-to-learn-Tableau-or-Power-BI)

Gambar 3.2 menunjukkan perbandingan kelebihan dan kekurangan dari masing-masing *tools*. Tableau memiliki keunggulan jika dibandingkan dengan *tools* lain. Dari gambar 3.2 hanya ada 2 hal yang tidak dimiliki oleh Tableau yakni *dashboard creation* dan *customize reporting*.

Perbandingan antara PowerBI dengan Tableau yakni ada 4 hal yang tidak dimiliki PowerBI jika dibandingkan dengan tableau yaitu *small business*, *phone*, *online*, dan *API*. Tapi ada beberapa hal yang tidak dimiliki oleh

PowerBI dan tidak dimiliki oleh Tableau yaitu *dashboard creation* yang dimana untuk *dashboard* adalah hal yang paling utama dalam analisa data karena *dashboard* adalah kunci utama dalam penyampaian pesan. Mudahnya dalam membaca *dashboard* tanpa harus dijelaskan terlebih dahulu akan sangat menentukan apakah hasil penelitian dapat tersampaikan dengan baik atau tidak.

Pentaho memiliki kekurangan yang lebih banyak jika dibandingkan dengan PowerBI dan tableau dimana pentaho memiliki 6 kekurangan yang tidak dimiliki oleh PowerBI. Perbedaan kekurangannya yang tidak dimiliki oleh *tools* lain yaitu pada bagian *knowledge base*, *video tutorial*, dan *data visualisation*. *Data visualisation* juga tidak kalah penting jika dibandingkan dengan *dashboard creation*. Kedua hal tersebut adalah ujung tombak atau hal utama dalam data visualisasi. Tanpa kedua hal tersebut tidak akan dapat menyampaikan informasi yang diinginkan. Selain itu pentaho menggunakan *command* atau sedikit menulis *code* sedangkan dalam PowerBI itu sendiri tidak perlu menggunakan *code* hanya tinggal menghitung biasa dan mengolahnya di komputer.

Penelitian ini menggunakan PowerBI karena sangat powerful dalam hal *dashboard* jika dibandingkan dengan *tools-tools* lain karena *dashboard* yang dapat dihasilkan oleh PowerBI sangat menarik dan mudah untuk dimengerti serta tidak membosankan untuk dilihat.

Akhirnya penelitian ini akan menggunakan PowerBI karena data-data yang ada yaitu data utang pemerintah dalam negeri Indonesia ini untuk memvisualisasikan dengan tepat dan interaktif, PowerBI menjadi *tools* yang paling tepat dan interaktif bagi pembaca agar mudah mengerti dan menangkap hasil yang disampaikan melalui visualisasi data utang pemerintah Indonesia ini.

